Reference: *Biol. Bull.* 189: 370–375. (December, 1995)

Interspecific Variations in Adhesive Protein Sequences of *Mytilus edulis, M. galloprovincialis,* and *M. trossulus*

KOJI INOUE[1,*], J. HERBERT WAITE[2], MAKOTO MATSUOKA[3], SATOSHI ODO[1], AND SHIGEAKI HARAYAMA[1]

[1]*Marine Biotechnology Institute, Kamaishi Laboratories, Heita, Kamaishi, Iwate 026, Japan;* [2]*Department of Chemistry and Biochemistry, University of Delaware, Newark, Delaware 19716; and* [3]*Juneau Center, School of Fishery and Ocean Sciences, University of Alaska, Fairbanks, Glacier Highway, Juneau, Alaska 99801*

Abstract. Variation in the adhesive protein gene sequences of *Mytilus edulis, M. galloprovincialis,* and *M. trossulus* collected in Delaware, Kamaishi (Japan), and Alaska, respectively, was analyzed by the polymerase chain reaction (PCR) using two sets of oligonucleotide primers. The first set, Me 13 and Me 14, was designed to amplify the repetitive region. The length of the amplified fragments was highly variable, even among samples of the same species. Another set, Me 15 and Me 16, was designed to amplify a part of the nonrepetitive region. The length of the amplified fragments was uniform in each species and differed interspecifically; 180, 168, and 126 bp for *M. edulis, M. trossulus,* and *M. galloprovincialis,* respectively. The amplified sequence of *M. trossulus* resembled that of *M. edulis.* Mussels from other sites were also examined by PCR using Me 15 and Me 16. Wild mussels from Tromsö (Norway) and cultured mussels from Brittany (France) were identified as *M. edulis.* Cultured mussels from the Mediterranean coast of France and wild mussels from Shimizu (Japan) were identified as *M. galloprovincialis.* Some wild mussels from Hiura (Japan) were identified as a hybrid between *M. galloprovincialis* and *M. trossulus.* Thus, the length of this part (variable region) of the sequence is proposed as a diagnostic marker for these three morphologically similar species and their hybrids.

Received 23 January 1995; accepted 10 August 1995.

The partial nucleotide sequence for *M. trossulus* adhesive protein gene reported in this paper has been submitted to the GenBank/EMBL/DDBJ Data Bank with accession number D50553.

* Address for correspondence: Koji Inoue at present address: Central Research Laboratory, Nippon Suisan Kaisha, Ltd., 559-6 Kitano, Hachioji, Tokyo 192, Japan.

Introduction

Two types of polyphenolic proteins, foot proteins 1 and 2, both of which incorporate 3,4-dihydroxyphenylalanine (DOPA) into their primary structures, have been isolated from the mussel *Mytilus edulis,* and characterized (see Waite, 1992, for a review). Foot protein 1 is an adhesive protein that contains repeats of the decapeptide motif AKPSYP*P*TY*K, where P* and Y* denote hydroxyproline and DOPA, respectively, and the hexapeptide motif AKPTY*K (Waite and Tanzer, 1981; Waite, 1983; Waite *et al.,* 1985). Foot protein 2 is an epidermal growth-factor-like protein that forms the adhesive plaque matrix (Rzepecki *et al.,* 1992; Inoue *et al.,* 1995a). The primary structure of the foot protein 1 has been determined by cDNA or gene cloning. It has been shown that it consists of a relatively short nonrepetitive domain and a long repetitive domain. The repetitive domain contains more than 70 decapeptide repeats and 13 or 14 hexapeptide repeats, but the number and distribution pattern of the motifs are variable even in the same species (Filpula *et al.,* 1990; Laursen, 1992).

We isolated cDNA encoding foot protein 1 from *M. galloprovincialis* sampled in Japan and compared the predicted amino acid sequence to those of *M. edulis* (Inoue and Odo, 1994; Inoue *et al.,* 1995b). The *M. galloprovincialis* sequence also consisted of the nonrepetitive and repetitive domains, but remarkable differences were observed in both. The major difference in the repetitive do-

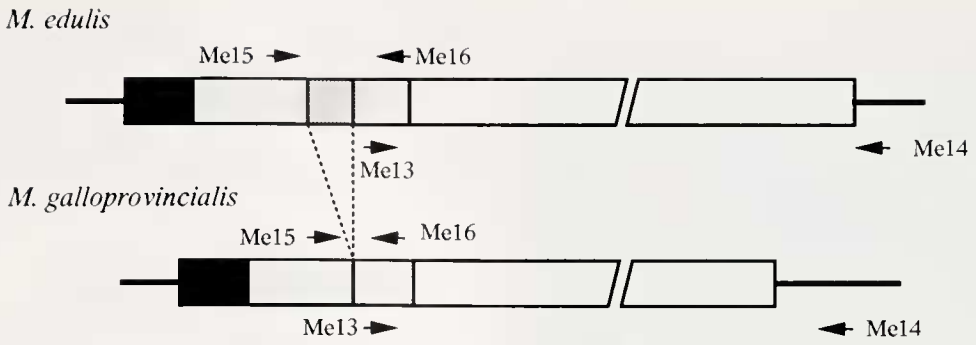

Figure 1. Positions of PCR primers in the adhesive protein genes of *Mytilis edulis* and *M. galloprovincialis*. Adhesive protein genes are shown as cDNA sequences. Position and direction of primers is indicated by arrows. Solid and open boxes indicate the signal peptide and the repetitive region, respectively. Dotted boxes indicate the nonrepetitive region, and the box with dense dots indicates the 18 amino acid sequence found in the *M. edulis* sequence but not in that of *M. galloprovincialis*. Solid lines indicate untranslated region.

main is that the *M. galloprovincialis* sequence contained 62 decapeptide repeats but no hexapeptide motif. In contrast, the nonrepetitive domains differ by a deletion of 18 amino acids observed in the *M. galloprovincialis* sequence.

In this study, we prepared two sets of oligonucleotide primers for polymerase chain reaction (PCR) to amplify the whole repetitive region and a part of the nonrepetitive region. Using these, we have analyzed foot protein 1 sequences of *M. edulis*, *M. galloprovincialis*, and *M. trossulus*. These three species are closely related (Gosling, 1984; Koehn, 1991; Gardner, 1992; Seed, 1992, for reviews) and practically indistinguishable by morphological characteristics only. We report that the length of the fragments amplified from the nonrepetitive region is specific to each species, but the length of the repetitive region is highly variable even within the same species. Thus the nonrepetitive region can be used as a diagnostic marker for identification of the three species.

Materials and Methods

Mussels

M. edulis was collected at Lewes (Delaware, USA); *M. galloprovincialis* and *M. trossulus* were sampled at Kamaishi (Iwate, Japan) and Juneau (Alaska, USA). All these sampling points are "pure sites" of each species where the other two species do not occur (McDonald *et al.*, 1991). Wild mussels were also collected at Tromsö (Norway), Hiura (Hokkaido, Japan), and Shimizu (Shizuoka, Japan). Mussels cultured in Brittany (France) and on the Mediterranean coast of France were obtained at a fish market at Ferney-Valtaire (France).

DNA extraction

A piece of the gill from each mussel, about 0.5 cm^2, was incubated in 500 μl lysis buffer containing 50 m*M* Tris-HCl (pH 7.5), 10 m*M* EDTA, 0.5% SDS, 500 μg/ml Proteinase K at 55°C for 2–4 h. Samples were then extracted twice with equal volumes of saturated phenol and twice with phenol chloroform: isoamyl alcohol (24:24:1). The aqueous phase was precipitated with ethanol and dissolved in 50 μl TE (10 m*M* Tris, 1 m*M* EDTA).

PCR amplification

About 100 ng of DNA was dissolved in 100 μl 1 × Tth buffer (TOYOBO, Japan) containing 6 μg sense primer, 6 μg antisense primer, and 200 μM dNTP. After preheating to 95°C, 1 unit of Tth DNA polymerase (TOYOBO, Japan) was added and 30 cycles of amplification were performed. Each cycle consisted of 30 s at 94°C, 30 s at 56°C, and 90 s at 70°C. The sequences of the primers were Me 13, CCA CTT GCA AAG AAG CTG TCA TCT; Me 14, ACA AAC GTT AAA ATG TGT AGT ACA GTA; Me 15, CCA GTA TAC AAA CCT GTG AAG A; Me 16, TGT TGT CTT AAT AGG TTT GTA AGA. Positions of primers in the foot protein 1 cDNA sequence in *M. galloprovincialis* are shown in Figure 1.

Electrophoresis of amplified products

Ten microliters of PCR product was mixed with the loading dye solution containing bromophenol blue (BPB) and xylene cyanol and subjected to agarose gel electrophoresis. For analysis of PCR products, 4.8% NuSieve 3:1 agarose (FMC) was used for the nonrepetitive region, and 1% LE agarose (FMC) was used for the repetitive region. Electrophoresis on a 4.8% gel was continued until BPB reached the end of the gel.

Sequencing

The fragment amplified from the genome of *M. trossulus* using primers Me 15 and Me 16 was isolated and

Figure 2. Representative results of amplification of the repetitive region of the adhesive protein gene. Amplified products were electrophoresed on 1% agarose gel. Lanes 1 and 2, *Mytilis edulis;* lanes 3 and 4, *M. trossulus;* lanes 5 and 6, *M. galloprovincialis.* M, molecular marker (lambda DNA digested with *Eco*T 141).

inserted into the *Sma* I site of pUC19. Sequences of both strands of three independent clones were determined using a 373A DNA sequencer (Applied Biosystems Inc.) and a PRISM Dyeterminator Cycle Sequencing Kit (Applied Biosystems Inc.).

Results

Variation in the repetitive region

The primers Me 13 and Me 14 were designed to amplify the repetitive region using the sequences identical to both *M. edulis* and *M. galloprovincialis.* Since the sense primer, Me 13, corresponds to a part of the nonrepetitive region and the antisense primer, Me 14, to a part of the 3′ untranslated region, the whole repetitive region is amplified by PCR. *M. edulis, M. galloprovincialis,* and *M. trossulus* were collected at Delaware, Kamaishi, and Juneau, respectively. These sampling points are known to be "pure sites" at which no other species of the *M. edulis* complex is found (McDonald *et al.,* 1991). We analyzed 8, 16, and 8 individuals of *M. edulis, M. galloprovincialis,* and *M. trossulus,* respectively, using primers Me 13 and Me 14. Since the repetitive region is relatively long and highly repetitive, it was difficult to amplify the whole repetitive region if the template DNA was insufficiently pure and long, but prominent bands were successfully obtained by using well-purified, high molecular weight DNA. Typical results are shown in Figure 2. Sizes of the band ranged from 2.2 to 2.8 kb. The fragments obtained from *M. edulis* were generally larger than those of the other two species. The sizes of bands in *M. trossulus* and *M. galloprovincialis* were similar but, on average, the former were slightly larger. Many individuals had two-banded (heterozygous) patterns, as expected for a highly variable polymorphism. One sample of *M. galloprovincialis* exhibited three bands, which may be a naturally occurring triploid or a mosaic individual that possesses a cell lineage having the differed length of foot protein 1 gene. It is, however, also possible that the third band is a heteroduplex of two different fragments.

Variation in the nonrepetitive region

Another set of primers, Me 15 and Me 16, was also prepared to amplify a part of the nonrepetitive region using sequences perfectly identical between *M. edulis* and *M. galloprovincialis* (Fig. 1). The size of the amplified fragment estimated from sequence data previously reported (Filpula *et al.,* 1990) is 180 bp in *M. edulis.* In *M. galloprovincialis,* the expected size is 126 bp because the sequence of *M. galloprovincialis* contains a deletion of 18 amino acids (Fig. 1; see also Inoue and Odo, 1994). Using these primers, 8, 32, and 16 individuals of *M. edulis, M. galloprovincialis,* and *M. trossulus*—including the same samples used in the analysis of the nonrepetitive region—were examined. PCR analysis indicated that all samples exhibited a single band. Representative results are shown in Figure 3. The position of the band was uniform in each species but differed from species to species. The size of the amplified fragments of *M. edulis* and *M. galloprovincialis* estimated by mobility in agarose agreed with those expected. Fragments from *M. trossulus* were shorter than those of *M. edulis* but longer than those of *M. galloprovincialis.* To determine the length and sequence of the amplified fragment of *M. trossulus,* the band obtained from one sample (Fig. 3, Lane 3) was isolated and se-

Figure 3. Representative results of amplification of the nonrepetitive region of the adhesive protein gene. Amplified products were electrophoresed on 4.8% NuSieve GTG agarose gel (FMC). Lanes 1 and 2, *Mytilis edulis;* lanes 3 and 4, *M. trossulus;* lanes 5 and 6, *M. galloprovincialis.* M, molecular marker (pUC19 DNA digested with *Hap*II).

```
        10        20        30        40        50        60
CCAGTATACAAACCTGTGAAGACAAGTTATTCGTCACCATATAAACCACCAACATACCAA
P  V  Y  K  P  V  K  T  S  Y  S  S  P  Y  K  P  P  T  Y  Q

        70        80        90       100       110       120
CCACTCAAAAAGAAACCGATGGACTATAATAGTTCTCCGCCAACATATGGATCAAAGACA
P  L  K  K  K  P  M  D  Y  N  S  S  P  P  T  Y  G  S  K  T

       130       140       150       160
AACTATCTTGCAAAGAAGCTGTCATCTTACAAACCTATTAAGACAACA
N  Y  L  A  K  K  L  S  S  Y  K  P  I  K  T  T
```

Figure 4. Nucleotide and deduced amino acid sequences of the fragment amplified from the genomic DNA of *Mytilis trossulus* using primers Me15 and Me16. Underlined sequences were derived from primers.

quenced. The sequenced fragment including the primer sequences was 168 bp, 12 bp shorter than the corresponding region of *M. edulis* (Fig. 4). This difference is small but measurable by mini-electrophoresis, as shown in Figure 3. The nucleotide sequence of the amplified fragment was compared with corresponding sequences of *M. edulis* and *M. galloprovincialis* (Fig. 5). Since the region shown in Figure 5 was especially variable among the three species, this region is hereafter referred to as the "variable region." It seems that variation among the three species was caused by deletion or addition of short sequences, not by base substitutions.

The variable region of mussels cultured in Brittany, those cultured on the Mediterranean coast of France, and the wild mussels collected at Tromsö and Hiura were also examined by PCR using Me 15 and Me 16. Four individuals were examined in each group and representative results are shown in Figure 6. All the wild Tromsö mussels and all the cultured Brittany mussels exhibited the 180-bp fragment and were identified as *M. edulis*. All the cultured Mediterranean mussels and the wild mussels from Shimizu exhibited a 126-bp fragment and were identified as *M. galloprovincialis*. These results are consistent with the distribution map of mussels that was made by using allozyme characters (McDonald *et al.*, 1991). Eight wild mussels of Hiura were also examined using Me 15 and Me 16. Two bands, 126 and 168 bp, were amplified from six individuals (Fig. 6), but only the 126-bp band was amplified from the remainder (data not shown). Hiura is on Hokkaido Island, where both native *M. trossulus* and introduced *M. galloprovincialis* are distributed. The mussels that

```
Mt: CAAGTTATTCGTCACCATATAAACCACCAACATACCAACCACTCAAAAAG
Me: **********G*************************************
Mg: ********-------------------------------------------

Mt: AAACCGATGGACTATAATAGT---------TCTCCGCCAACATATGGATC
Me: ***---G********CG*CC*ACGAAAAGT*A********************
Mg: ---------------CA*CC*ACGAATAGT*A********************

Mt: AAAGACAAACTAT------CTTGCAAAGAAGCTGTCA
Me: **************CTACCA******************
Mg: **************CTGCCA******************
```

Figure 5. Comparison of nucleotide sequences of the variable region of the adhesive protein genes of *Mytilis trossulus*, *M. edulis*, and *M. galloprovincialis*. Sequences of *M. edulis* and *M. galloprovincialis* were according to Filpula *et al.* (1990) and Inoue and Odo (1994), respectively. Asterisks indicate nucleotides identical to those of the *M. trossulus* sequence. Hyphens indicate gaps inserted to align sequences. Mt, *M. trossulus*; Me, *M. edulis*; Mg, *M. galloprovincialis*.

Figure 6. Amplification of the variable region of the adhesive protein gene of wild and cultured mussels. Amplified products were electrophoresed on 4.8% NuSieve GTG agarose gel (FMC). Lanes 1–3, wild mussels collected at Tromsö; Lanes 4–6, mussels cultured in Brittany; Lanes 7–9, wild mussels collected at Hiura; Lanes 10–12, mussels cultured on the Mediterranean coast of France. M, molecular marker (pUC19 DNA digested with *Hap*II).

exhibited two fragments are presumed to be hybrids between the two species.

Discussion

Among the five species of the genus *Mytilus*, *M. edulis*, *M. galloprovincialis*, and *M. trossulus* have been called the "*M. edulis complex*." Since they are morphologically similar and shell shape is often influenced by local environment, it is difficult to identify these species by morphological characteristics. Recently, allozyme characters have been used to clarify the taxonomy of these species (Koehn *et al.*, 1984; McDonald and Koehn, 1988; Varvio *et al.*, 1988; McDonald *et al.*, 1991; Coustau *et al.*, 1991; Viard *et al.*, 1994). These characters are recognized as reliable markers, but data for multiple loci are required for accurate identification of all three species. Identification using mitochondrial DNA (mtDNA) sequences has also been described (Edwards and Skibinski, 1987; Blot *et al.*, 1990; Geller *et al.*, 1993, 1994). Although such attempts were partially successful, it is still difficult to differentiate the three species unambiguously. In this study, we found that differences in a certain "variable region" of a sequence in the nonrepetitive domain of the foot protein 1 agree well with the taxonomic rank of species. It was also shown that the variations can be attributed to differences in the length of the fragments amplified by PCR. Thus the variable region may become an effective diagnostic marker. Because PCR requires only a small amount of DNA as a template, the method may be used for larvae or young individuals that are too small for analysis by other methods. In addition, the PCR system seems effective for the identification of hybrids within the *M. edulis* complex—we could detect hybrids that have the haplotypes of both *M. galloprovincialis* and *M. trossulus* (Fig. 6). This system may become a powerful tool for studying the distribution and genetics of mussels, one of the most cosmopolitan of marine animals.

We also tried to amplify the adhesive protein gene of *M. coruscus*, a mussel species native to Japan, but we were unsuccessful (data not shown). The adhesive protein gene sequence of *M. coruscus* may be considerably different from those of species in the *M. edulis* complex. It seems that other primers are required for analysis of mussels other than the *M. edulis* complex.

In contrast to the nonrepetitive domain, the length of the repetitive domain was highly variable, even in samples collected at the same site. This result is consistent with the fact that repeat patterns observed in three nucleotide sequences encoding foot protein 1 of *M. edulis* reported separately (Strausberg *et al.*, 1989; Filpula *et al.*, 1990; Laursen, 1992) differ from one another. Because the length of this region was partially overlapping among species, it seems inappropriate for use as a marker of species. Many individuals of the three species also have two or more fragments of different length. These results suggest that each species has enormous variability in the patterns of repetitive sequences. Considering that foot protein 1 is a key molecule for adhesion of mussels and thus is essential for their survival, its extensive diversity is intriguing. We suggest that the repetitive domain has been differentiated by a "shuffling" of the repeat pattern as well as by the common base substitution process (Inoue *et al.*, 1995b). Increased diversity may have an important role in mussel survival or evolution.

Acknowledgments

The authors express their sincere thanks to Dr. Shigetoh Miyachi for support in this study, to Drs. Shigeru Nakao and Takashi Noda for valuable advice, to Hiroyuki Kawahara for collecting mussels, and to Sachiko Dobashi for technical assistance. This work was performed as a part of the Industrial Science and Technology Frontier Program supported by New Energy and Industrial Technology Development Organization.

Literature Cited

Blot, M., B. Legendre, and P. Albert. 1990. Restriction fragment length polymorphism of mitochondrial DNA in subantarctic mussels. *J. Exp. Mar. Biol. Ecol.* **141**: 79–86.

Coustau, C., F. Renaud, and B. Delay. 1991. Genetic characterization of the hybridization between *Mytilus edulis* and *M. galloprovincialis* on the Atlantic coast of France. *Mar. Biol.* **111**: 87–93.

Edwards, C. A., and D. O. F. Skibinski. 1987. Genetic variation of mitochondrial DNA in mussel (*Mytilus edulis* and *M. galloprovincialis*) population from South West England and South Wales. *Mar. Biol.* **94**: 547–556.

Filpula, D. R., S-M. Lee, R. P. Link, S. L. Strausberg, and R. L. Strausberg. 1990. Structural and functional repetition in a marine mussel adhesive protein. *Biotechnol. Prog.* **6**: 171–177.

Gardner, J. P. A. 1992. *Mytilus galloprovincialis* (Lmk) (Bivalvia, Mollusca): the taxonomic status of the Mediterranean mussel. *Ophelia* **35:** 219–243.

Geller, J. B., J. T. Carlton, and D. A. Powers. 1993. Interspecific and intrapopulation variation in mitochondrial ribosomal DNA sequences of *Mytilus* spp. (Bivalvia: Mollusca). *Mol. Mar. Biol. Biotechnol.* **2:** 44–50.

Geller, J. B., J. T. Carlton, and D. A. Powers. 1994. PCR-based detection of mtDNA haplotypes of native and invading mussels on the northeastern Pacific coast: latitudinal pattern of invasion. *Mar. Biol.* **119:** 243–249.

Gosling, E. M. 1984. The systematic status of *Mytilus galloprovincialis* in western Europe: a review. *Malacologia* **25:** 551–568.

Inoue, K., and S. Odo. 1994. The adhesive protein cDNA of *Mytilus galloprovincialis* encodes decapeptide repeats but no hexapeptide motif. *Biol. Bull.* **186:** 349–355.

Inoue, K., Y. Takeuchi, D. Miki, and S. Odo. 1995a. Mussel adhesive plaque protein gene is a novel member of epidermal growth factor-like gene family. *J. Biol. Chem.* **270:** 6698–6701.

Inoue, K., Y. Takeuchi, D. Miki, and S. Odo. 1995b. Mussel adhesive protein genes: structure and variations. *J. Mar. Biotechnol.* (in press).

Koehn, R. K. 1991. The genetics and taxonomy of species in the genus *Mytilus*. *Aquaculture* **94:** 125–145.

Koehn, R. K., J. G. Hall, D. J. Innis, and A. J. Zera. 1984. Genetic differentiation in *Mytilus edulis* in eastern North America. *Mar. Biol.* **79:** 117–126.

Laursen, R. A. 1992. Reflections on the structure of mussel adhesive proteins. Pp. 55–74 in *Structure, Cellular Synthesis and Assembly of Biopolymers. Results and Problems in Cell Differentiation*, Vol. 19, S. T. Case, ed. Springer-Verlag, Berlin.

McDonald, J. H., and R. K. Koehn. 1988. The mussels *Mytilus galloprovincialis* and *M. trossulus* on the Pacific coast of North America. *Mar. Biol.* **99:** 111–118.

McDonald, J. H., R. Seed, and R. K. Koehn. 1991. Allozymes and morphometric characters of three species of *Mytilus* in the Northern and Southern Hemispheres. *Mar. Biol.* **111:** 323–333.

Rzepecki, L. M., K. M. Hansen, and J. H. Waite. 1992. Characterization of a cystine-rich polyphenolic protein family from the blue mussel *Mytilus edulis* L. *Biol. Bull.* **183:** 123–137.

Seed, R. 1992. Systematics evolution and distribution of mussels belonging to the genus *Mytilus*: an overview. *Am. Malac. Bull.* **9:** 123–137.

Strausberg, R. L., D. M. Anderson, D. Filpula, M. Finkelman, R. Link, R. McCandliss, S. A. Orndorff, S. L. Strausberg, and T. Wei. 1989. Development of a microbial system for production of mussel adhesive protein. Pp. 453–464 in *Adhesives from Renewable Resources*. ACS Symposium Series 385, R. W. Hemingway and A. H. Conner, eds. American Chemical Society, Washington, DC.

Varvio, S-L., R. K. Koehn, and R. Väinölä. 1988. Evolutionary genetics of the *Mytilus edulis* complex in the North Atlantic region. *Mar. Biol.* **98:** 51–60.

Viard, F., B. Delay, C. Coustau, and F. Renaud. 1994. Evolution of the genetic structure of bivalve cohorts at hybridization sites of the *Mytilus edulis–M. galloprovincialis* complex. *Mar. Biol.* **119:** 535–539.

Waite, J. H. 1983. Evidence for a repeating Dopa and hydroxyproline containing decapeptide in the adhesive protein of *Mytilus edulis*. *J. Biol. Chem.* **258:** 2911–2915.

Waite, J. H. 1992. The formation of mussel byssus: anatomy of a natural manufacturing process. Pp. 27–54 in *Structure, Cellular Synthesis and Assembly of Biopolymers. Results and Problems in Cell Differentiation*, Vol. 19, S. T. Case, ed. Springer-Verlag, Berlin.

Waite J. H., and M. L. Tanzer. 1981. Polyphenolic substance of *Mytilus edulis*. *Science* **212:** 1038–1040.

Waite, J. H., T. J. Housley, and M. L. Tanzer. 1985. Peptide repeats in a mussel glue protein: theme and variations. *Biochemistry* **24:** 5010–5014.